

Appendix A

History of the Toxicology Information Program of the National Library of Medicine

Background

The 1960s were an era of growing public environmental awareness. The publication in 1962 of Rachel Carson's *Silent Spring* (1) was a powerful spur to the American public and the Congress to take action to protect the environment. A 1966 report by the President's Science Advisory Committee (PSAC) outlined issues related to the effects of hazardous chemicals on the environment and human health. The report not only presented an early discussion on the significance of toxicology to society, it highlighted the importance of ready access to information.

The report suggested that the Department of Health, Education, and Welfare (now the Department of Health and Human Services) was the appropriate agency to develop an information program in this area. President Johnson accepted the recommendations of this report and, as a result, the Department directed the National Library of Medicine to establish a Toxicology Information Program.

At the start, Dr. Martin M. Cummings, then NLM Director, had anticipated receiving approximately 40 personnel positions with an annual budget of several million dollars for the Program. But after a few years it became apparent that TIP would be limited to less than half the positions and to about a million dollars annually. The plans had to be restricted and TIP, thereafter, had to proceed on a much smaller scale than had been hoped for.

In 1969, the Program started what was to become a long-lasting relationship with the National Academy of Sciences

for the establishment of an advisory committee—the Toxicology Information Program Committee (TIPCOM). It was intended that this group would be a channel through which the information needs and opinions of academic and industrial biomedical scientists could be brought to the attention of the TIP. Furthermore, all projects that the Program has undertaken have been passed in review before this group.

The scope of information covered by NLM's toxicology files has evolved over the years. Originally slanted more toward adverse health effects of selected drugs and chemicals, the files currently are influenced by a much broader and deeper assortment of societal concerns about chemicals and other contaminants. Some of this is mandated by recent environmental legislation, as shown by specific information-related mandates in such major laws as the Toxic Substances Control Act of 1976 and the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (the "Superfund" legislation) and its amended version of 1986.

Such mandates for new, environmentally related information activities and services had major impacts on the scope and nature of NLM's service. In many instances, the agency charged in the legislation with implementation of such activities started to collaborate with NLM and built the new activities by enhancing existing NLM services. For example, ATSDR now provides substantial support to NLM for the maintenance of the Hazardous Substances Data Bank (HSDB) which furnishes detailed information on the properties, emergency handling, environmental fate and regulatory requirements for over 4,200 hazardous chemicals.

The substantial contribution of funding from other agencies in support of projects of mutual interest has been helpful. In fact, certain projects could not have been started or sustained without such support. However, there is a downside to obtaining basic funding support for projects from others, because it ties the fate of the supported projects to the fortunes of the other agency. It would be preferable if funds for such collaborative projects could be shifted to NLM at the Congressional appropriations level and thus become part of the NLM budget.

Information Services

Reference Services

From its inception, TIP has provided references services in response to queries from the biomedical community. However, because internal staff resources were insufficient for a full-fledged service, an interagency agreement with the Atomic Energy Commission (now the Department of Energy) was established and used to set up in 1972 a Toxicology Information Response Center (TIRC) at the Oak Ridge National Laboratory (ORNL). TIRC searched NLM databases and other sources, and furnished comprehensive literature searches to requestors on a fee-for-service basis.

Over the years, the increasing availability and popularity of online retrieval services in toxicology has greatly diminished the demand for the kind of specialized reference service provided by TIRC in the 70s. The Center is still in operation but its size has been reduced; it now provides responses to some 15 search requests per month, and many of these come from the ORNL science community itself.

Publications

While the main method by which the TIP now provides information to its clients is via its online computer services, initially the Program put a certain amount of effort into publishing bibliographies, monographs and periodicals. Among the titles published over the years (some in collaboration with other organizations) were: *Drug Interactions: An Annotated Bibliography with Selected Excerpts* (1967-1971); *Toxicity Bibliography* (1968-78); *A Directory of Information Resources in the United States: General Toxicology* (1969); *Pesticides Abstracts*; *Abstracts on Health Effects on Environmental Pollutants*; *Toxicology Research Projects Directory*; *Tox-Tips (Toxicology Testing in Progress)* (1976-87); *Symposium on the Handling of Toxicological Information* (1976); *NTP Chemical Registry Handbook* (1981); *Proceedings of the Symposium on Information Transfer in Toxicology* (1982); *Alternatives to the Use of Live Vertebrates in Biomedical Research and Testing: An Annotated Bibliography* (1988).

Online Bibliographic Services

TOXLINE, the first online bibliographic system for toxicology, was developed by the TIP in 1972 as a "one-stop shopping center" for toxicology information. The original intent to follow the MEDLINE lead and "mechanize" an existing abstracting and indexing (A&I) source for online bibliographic retrieval had to be adjusted because no one secondary source covered the field of toxicology sufficiently. It was decided, therefore, to combine "toxicology subsets" from various A&I services into one file that would look reasonably homogeneous to the online user. Thus, TOXLINE initially incorporated relevant segments from

Index Medicus, *Biological Abstracts*, *Chemical Abstracts*, and *International Pharmaceutical Abstracts*. Over the years other segments have been added while some had to be deleted. TOXLINE also served to validate the utility of whole-text searching without a controlled vocabulary. This was accomplished by creating one large inverted file of all searchable terms.

Several of the TOXLINE segments were obtained from A&I services under usage royalty agreements. The costs were passed on to the users and, over time, increased to where TOXLINE became substantially more expensive than other NLM online services. License arrangements, not involving usage royalties, could be developed for some segments, notably those from *Biological Abstracts* and *International Pharmaceutical Abstracts*. However this did not prove feasible with *Chemical Abstracts*. In 1987, it was decided to split TOXLINE into a new usage royalty service called TOXLIT (and its backfile TOXLIT65), which contained the *Chemical Abstracts*-derived segments, and another service, without such royalties, for which the name of TOXLINE was kept. This family of services now contains more than 3 million records.

Online Chemical Dictionaries and Directories

Because toxicology is concerned with the effects of chemicals on biological systems, the accurate identification of chemical substances is often a critical preliminary to utilizing toxicology information systems. For TOXLINE, this problem was met by building CHEMLINE, an online chemical dictionary file that derived its content mainly from the Chemical Abstracts Service (CAS)

Registry System. CHEMLINE was the forerunner of the class of online chemical information directories in that it linked nomenclature, structural information and CAS Registry Numbers to the location of information about a specific chemical or groups of structurally related chemicals in other files. CHEMLINE made two fundamental contributions to chemical information retrieval: it demonstrated the importance of the CAS Registry Number in online information seeking, and it showed that fragments derived from parsing standardized chemical nomenclature could provide useful online substructure searching capabilities.

Over the years, as more NLM files with chemical subject content became available, CHEMLINE expanded its directory function by pointing users to these other NLM files where information about a chemical or a family of chemicals might be found. To obtain the data needed to build CHEMLINE, an agreement involving usage royalties was reached with CAS. It gave NLM access to certain computer-produced portions of the CAS Registry System. As was the case for TOXLINE, NLM adopted a policy whereby the royalty charges arising from the use of CAS information were passed on to the users of CHEMLINE. Because CAS has steadily raised those royalty charges, TIP developed a new, royalty-free chemical information resource named ChemID (**C**hemical **I**dentification) that now contains some 200,000 records. ChemID was created using a variety of royalty-free sources. It still functions as a chemical dictionary with nomenclature, synonyms, and CAS Registry Numbers, and as a directory with pointers to the location of information about chemicals in other MEDLARS

files, with the exception of the royalty-based TOXLIT and TOXLIT65 files. ChemID was made available to the public in February 1990.

In July 1990, ChemID's directory functions were substantially expanded with a new set of data collectively known as SUPERLIST. This capability provides the names and other chemical identifiers used on a variety of regulatory or scientific lists. With the proliferation of Federal, state, and international lists of chemicals, the presence of a chemical on such a list can have important regulatory or financial consequences for government agencies as well as for commercial organizations. At the present time, this coverage includes 20 lists such as the Department of Transportation Hazardous Materials List, EPA Pesticide List, OSHA Toxic and Hazardous Substances, and the IARC (International Agency for Research on Cancer) list.

Online Data Banks

Bibliographic retrieval systems—online or hard copy—are fact locators in that they direct the user to journal articles or books that contain the sought-for facts. In contrast, data or fact retrieval systems—like handbooks—provide users with the actual facts.

Having organized TOXLINE, the bibliographic retrieval service for toxicology, the Toxicology Information Program turned, in 1974, to the creation of the Toxicology Data Bank (TDB), an online-searchable compilation of referenced numeric data and textual statements about the chemical and biological properties of some 3000 drugs and hazardous chemicals. Data statements were extracted from published sources such as

textbooks, monographs, Government reports and, to a lesser extent, journal articles. The file was organized by compound, and records were reviewed periodically for validity and accuracy by a special review group—the Scientific Review Panel (SRP). The “grant-like” review process of reviewing TDB records is a unique way of reviewing the content of scientific documents by a committee. The TDB was later renamed the Hazardous Substances Data Bank (HSDB).

TOXNET

Initially, the TDB was maintained and provided to the public through ELHILL, NLM's major file building and information retrieval system. However, that system was originally designed for efficient bibliographic retrieval from large files with many, relatively small records. It never functioned well for the kind of data retrieval required for the TDB. When it became clear that ELHILL would not be able to handle the large data records resulting from the expansion of the TDB to the HSDB, the Program developed a new system, named TOXNET.

TOXNET is a multicomponent system that allows for online file building, content review and revisions, as well as online, interactive searching. Access to TOXNET is available to all NLM users. Through a gateway that links the TOXNET computers to the NLM mainframe, TOXNET users also can readily access relevant NLM bibliographic files such as MEDLINE or TOXLINE. Online usage of the TOXNET files continues to grow and now averages 1600 hours per month; another 250 hours are used for gateway access to the files on the NLM mainframe.

The following files (and their sponsoring organizations) are available on TOXNET:

Chemical Carcinogenesis Research Information System (CCRIS), NCI

Developmental and Reproductive Toxicology (DART), EPA & NIEHS

Environmental Mutagen Information Center (EMIC), EPA & NIEHS

EMICBACK (Backfile for EMIC)

ETICBACK (Backfile for DART)

Genetic Toxicology (GENETOX), EPA

Hazardous Substances Data Bank (HSDB), NLM & ATSDR

Integrated Risk Information System (IRIS), EPA

Registry of Toxic Effects of Chemical Substances (RTECS), NIOSH

Toxic Chemical Release Inventory for 1987, 1988, 1989, 1990, EPA

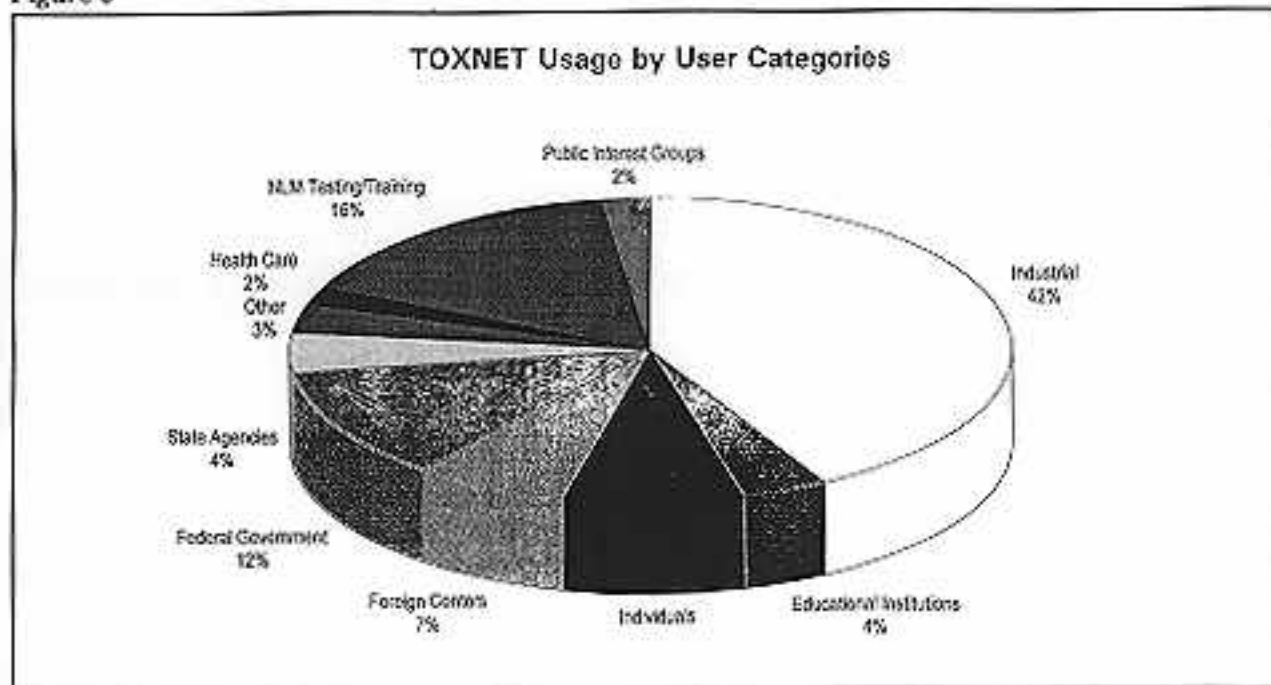
While most TOXNET files take their content from published scientific

sources, one family of files—EPA's Toxic Chemical Release Inventory (TRI)—contains data on chemical releases reported by industry to the EPA as mandated by the SARA legislation. This set of files includes substantial amounts of numeric data, and for it, TOXNET had to be enhanced with statistical, ranging, sorting and menu-based searching capabilities.

Users

When a user organization applies for an access code to the NLM online services, its affiliation is recorded (e.g., industrial, Federal government, state government). Usage is then tracked for billing purposes, and it is relatively easy to link usage by file to user classes. Applying some recent use figures for the files on the TOXNET system as indicative of the usage of the Library's toxicology files in general shows the distribution diagrammed in Figure 6.

Figure 6



Because the Library does not normally examine the computer interactions involved in a user's search session, it is more difficult to obtain reliable data concerning the purposes for which the NLM online services are being used. In order to gather some information about what users are expecting to garner from the toxicology data files on TOXNET, a preliminary user needs assessment was carried out in July 1991. Toxicity and health effects led the list of types of information sought by those responding to the survey. A large number of respondents also indicated a need for the types of information found in the TRI (Toxic Chemical Release Inventory) files.

Systems Redesign

As TOXNET grew in size and usage, it became clear that NLM needed to consider some fundamental changes in the basic design of the system. Because the TOXNET MUIS language/operating system functioned primarily on Data General minicomputers, TOXNET was tied to use of this equipment. Equipment costs were high and future support for this line of hardware was cloudy.

An evaluation study of TOXNET, carried out in 1989 by a team of outside computer scientists, recommended that NLM rewrite the system in a variant of the MUMPS language, which is available to operate on many hardware platforms. This recommendation was followed, and TOXNET now operates in standard ANSI MUMPS on a cluster of microcomputers. This state-of-the-art configuration is more efficient and less costly than the previous system.

In parallel with this effort, it was also decided to explore the use of Relational Database Management System (RDBMS) technology for the family of files now on TOXNET. True file integration would result in more efficient file management, and allow various approaches to the data supporting a variety of user purposes. Conceptually, TOXNET lends itself well to such an approach because it contains multiple files with related—sometimes duplicative—information.

In June 1991, NLM convened a panel of computer experts to address the question of whether relational technology is suitable for these databases. This panel recommended that NLM proceed with developing an RDBMS-based file building module for the HSDB and some of the other TOXNET files but, for the time being, stay away from converting the entire TOXNET system to a relational system, because none of the existing RDBMS packages are efficient enough in online searching of highly textual files such as many of those on TOXNET.

Emergency Response

The Superfund Act included requirements for support of emergency response activities in cases of accidents involving hazardous chemicals. In response to these mandates, ATSDR collaborated with NLM, to build a transportable, microcomputer-based workstation that provides information assistance to emergency response teams working on such accidents. The operational prototype, known as ANSWER (an acronym for ATSDR-NLM's Workstation for

Emergency Response), consists of software modules designed to facilitate easy access to information useful to response teams during emergencies.

The core modules of the Workstation are: a CD-ROM containing databases with information on both hazard management and medical management; a special database with information gleaned from previous chemical emergencies; a modified version of software (Micro-CSIN) that facilitates searching of diverse, remote online databases; a FAX capability to transmit information to and from an emergency site; access to weather information from the National Weather Service; and a word processing capability for dealing with retrieved data.

The Workstation was made available for Beta testing at 13 sites, including selected state health departments and several poison control centers. The results of the test showed that ANSWER is highly useful in a command center environment in both emergency and nonemergency situations, and that additional chemical databases on CD-ROM would be very helpful in the field.

Issues

Several issues of potential significance for the future of the Toxicology Information Program can be extracted from its 24-year history:

When taking inflation into account, the NLM portion of the TIP budget has remained constant since 1968. However, several new products were developed with funds from collaborating agencies and with funds collected from users. Funding through other agencies is

intrinsically unstable because it depends on their budgets and program objectives.

Leasing files from organizations that require usage royalties was useful because it brought valuable content to the NLM files. However, the steady increase in these royalties made it necessary to restructure the NLM files into royalty and nonroyalty services so that users that could not afford the higher charges, could still make use of a portion of the information.

In part, NLM evaluates the utility of its online files by the amount of usage. In recent years, several of these files have been leased from NLM by the vendors of CD-ROM products for incorporation in commercial CD-ROM products. As these products become more widely accepted, usage of NLM's online files will decrease without NLM being able to measure the offsetting uses of the commercial CD-ROM products.

Generally, NLM services are intended for use by scientists and health professionals. However, the Toxic Chemical Release Inventory (TRI) files on TOXNET are clearly meant by law to be available and accessible to the general population. This altered focus placed new requirements for ease of access and simplicity of content on the Program.

Searching NLM's online toxicology files is difficult. This has put a substantial training burden on the Program. In the future, such training costs should be included when considering the cost-effectiveness of online systems. In the long run, a more expensive, easy-to-use system may be cheaper than a complex system requiring training courses and teaching packages.

Conclusion

The 25 years since issuance of the PSAC report have seen a steadily increasing concern, worldwide, with the threats to human welfare stemming from environmental contamination. Widely available, reliable scientific and technical information is an important tool in the campaign to clean up the environment and protect human health. The Toxicology Information Program has been an active participant in this campaign for a quarter century and is ready to embrace a continuing and expanded role in the future.

References

- (1) Carson, R. *Silent Spring*. Boston: Houghton Mifflin; 1962.
- (2) President's Science Advisory Committee. *Handling of Toxicological Information*. Washington, D.C.: The White House, 1966.
- (3) Miles, WD. Chapter XII, Specialized Information Services. *A History of the National Library of Medicine: the Nation's Treasury of Medical Knowledge*. Bethesda, MD: National Library of Medicine, 1982; NIH publication no. 82-1904.
- (4) Kissman, HM., Wexler, P. Toxicology Information Systems: A Historical Perspective. *J Chem Inf Comp Sci* 1985; 25: 212-17.
- (5) Wexler, Philip. Chapter 10, Legislation and Regulatory Issues. *Information Resources in Toxicology*. New York: Elsevier, 1988; 245-253.